

Article Type: Special Issue Article

RESEARCH ARTICLE

INVITED SPECIAL ARTICLE

For the Special Issue: Using and Navigating the Plant Tree of Life

Short Title: Pease et al.—Quartet Sampling of discordance in the green plant tree of life

**Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life**

James B. Pease<sup>1,4</sup>, Joseph W. Brown<sup>2</sup>, Joseph F. Walker<sup>2</sup>, Cody E. Hinchliff<sup>3</sup>, and Stephen A. Smith<sup>2,4</sup>

Manuscript received 9 June 2017; revision accepted 5 September 2017.

<sup>1</sup> Department of Biology, Wake Forest University, 455 Vine Street, Winston-Salem, North Carolina 27101 USA

<sup>2</sup> Department of Ecology and Evolutionary Biology, University of Michigan, 830 North University, Ann Arbor, Michigan 48109 USA

<sup>3</sup> Department of Biological Sciences, University of Idaho, 875 Perimeter Drive, MS 3051, Moscow, Idaho 83844 USA

<sup>4</sup> Authors for correspondence (e-mails: [peasejb@wfu.edu](mailto:peasejb@wfu.edu), [eebsmith@umich.edu](mailto:eebsmith@umich.edu))

**Citation:** Pease, J. B., J. W. Brown, J. F. Walker, C. E. Hinchliff, and S. A. Smith. 2018. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *American Journal of Botany* 105(3): XXX.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/ajb2.1016](https://doi.org/10.1002/ajb2.1016)

This article is protected by copyright. All rights reserved

**DOI: XXXX**

**PREMISE OF THE STUDY:** Phylogenetic support has been difficult to evaluate within the green plant tree of life partly due to a lack of specificity between conflicted versus poorly informed branches. As data sets continue to expand in both breadth and depth, new support measures are needed that are more efficient and informative.

**METHODS:** We describe the Quartet Sampling (QS) method, a quartet-based evaluation system that synthesizes several phylogenetic and genomic analytical approaches. QS characterizes discordance in large-sparse and genome-wide data sets, overcoming issues of alignment sparsity and distinguishing strong conflict from weak support. We tested QS with simulations and recent plant phylogenies inferred from variously sized data sets.

**KEY RESULTS:** QS scores demonstrated convergence with increasing replicates and were not strongly affected by branch depth. Patterns of QS support from different phylogenies led to a coherent understanding of ancestral branches defining key disagreements, including the relationships of *Ginkgo* to cycads, magnoliids to monocots and eudicots, and mosses to liverworts. The relationships of ANA-grade angiosperms (*Amborella*, Nymphaeales, Austrobaileyales), major monocot groups, bryophytes, and fern families are likely highly discordant in their evolutionary histories, rather than poorly informed. QS can also detect discordance due to introgression in phylogenomic data.

**CONCLUSIONS:** Quartet Sampling is an efficient synthesis of phylogenetic tests that offers more comprehensive and specific information on branch support than conventional measures. The QS method corroborates growing evidence that phylogenomic investigations that incorporate discordance testing are warranted when reconstructing complex evolutionary histories, in particular those surrounding ANA-grade, monocots, and nonvascular plants.

**KEY WORDS:** bootstrap; branch support; discordance; introgression; likelihood; lineage sorting; phylogenetics; phylogenomics; *Viridiplantae*

Discordance and uncertainty have emerged as consistent features throughout the history of our evolving model of the green plant tree of life (Crane, 1985; Chase et al., 1993; Palmer et al., 2004; Soltis et al., 2011; Wickett et al., 2014). Particularly strong contentions often arise at pivotal transitions in the evolution of plant life on earth, such as the development of vascular tissue (Pryer et al., 2001; Steemans et al., 2009; Banks et al., 2011), the rise of seed-bearing plants (Chase et al., 1993; Chaw et al., 1997; Bowe et al., 2000; Qiu et al., 2006; Jiao et al., 2011), and the explosive radiation of flowering plants (Crane, 1985; Amborella Genome Project, 2013; Goremykin et al., 2015; Taylor et al., 2015; Edwards et al., 2016; Simmons, 2016). Modern phylogenomic data sets, rather than quelling these disagreements, have repeatedly shown that these phylogenetic conflicts are often the result of biological processes including incomplete lineage sorting (ILS), introgressive hybridization, and paralog duplication and loss (e.g., Zhong et al., 2013b; Wickett et al., 2014; Zwickl et al., 2014; Yang et al., 2015; Pease et al., 2016b; Eaton et al., 2017; Goulet et al., 2017; Walker et al., 2017c). Several methods have been proposed to address these issues during species tree inference (e.g., Zwickl and Hillis, 2002; Ogden and Rosenberg, 2006; Shavit Grievink et al., 2010; Anderson et al., 2012; Roure et al., 2012; Hinchliff and Roalson, 2013; Mirarab et al., 2014). However, we lack a generalized framework to quantify phylogenetic uncertainty (specifically branch support) that distinguishes branches with low information from those with multiple highly supported, but mutually exclusive, phylogenetic histories.

One of the most commonly used branch support methods has been the nonparametric bootstrap (NBS; Felsenstein, 1985) and recent variants such as the rapid bootstrap (RBS; Stamatakis et al., 2008), which resample the original data with replacement assuming that aligned sites are independent and identically distributed (i.i.d.) samples that approximate the true underlying distribution (Felsenstein, 1985; Efron, 1992). In practice, the assumptions of NBS (in particular site independence) may rarely be met and can deteriorate under a variety of conditions (Felsenstein and Kishino, 1993; Hillis and Bull, 1993; Sanderson, 1995; Andrews, 2000; Alfaro et al., 2003; Cummings et al., 2003). More recently, the UltraFast bootstrap approximation (UFboot) method proposed likelihood-based candidate tree testing to address speed and score interpretation issues for NBS (Minh et al., 2013; see comparison by Simmons and Norton, 2014).

The other commonly used branch support metric has been the Bayesian posterior probability (PP). PP scores are typically calculated from posterior distributions of trees generated

using a Markov chain Monte Carlo (MCMC) sampler and summarized using a majority-rule consensus tree (e.g., Larget and Simon, 1999; Drummond and Rambaut, 2007; Holder et al., 2008; Ronquist et al., 2012; Larget, 2013). The interpretation of PP values is more straightforward than bootstrap proportions because PP values represent the probability that a clade exists in the underlying tree, conditioned on the model of evolution employed and the prior probabilities. The individual and relative performance of PP has been well documented as generally favorable (Wilcox et al., 2002; Alfaro et al., 2003; Cummings et al., 2003; Huelsenbeck and Rannala, 2004). However, PP may be excessively high in certain scenarios (e.g., oversimplified substitution models; Suzuki et al., 2002; Douady et al., 2003; Nylander et al., 2004). Bayesian posterior probability also may fail under a multispecies coalescent framework with conflicting phylogenies (Reid et al., 2013), which is particularly noteworthy in light of studies showing the disproportionate effects of a few genes on overall genomewide phylogenies (Brown and Thomson, 2017; Shen et al., 2017; Walker et al., 2017a).

Ongoing efforts to expand genetic sampling to as many plant species as possible have produced increasingly species-rich, but data-sparse, alignments (i.e., large-sparse or “fenestrated” matrices). Meanwhile, the accelerating accretion of new genomes and transcriptomes will continue to deepen genome-wide data sets with millions of aligned sites. Both axes of data set expansion present challenges to the tractability and interpretation of phylogenetic branch-support analytics. Nonparametric bootstrap scores are known to perform poorly for large-sparse matrices when the sampling procedure generates uninformative pseudoreplicates that mostly omit informative sites or consist of mostly missing data (Driskell et al., 2004; Smith et al., 2011; Wiens and Morrill, 2011; Roure et al., 2012; Hinchliff and Roalson, 2013; Hinchliff and Smith, 2014b). Furthermore, resampling methods (including NBS) approximate the resampling of a larger idealized population. Genomic data sets contain virtually all available data and, therefore, are not samples of any larger whole. Bayesian posterior probabilities provide an appropriate testing framework and straightforward interpretation for genomic data, but available Bayesian methods of analysis are not scalable to genome-wide data under current computational speeds (though see tools like ExaBayes; Aberer et al., 2014). They may also may over-estimate support when models are overly simple, which becomes increasingly problematic with expanding size and complexity of data sets. Posterior probability and NBS scores therefore both appear unsuitable for use on

large data sets due to feasibility and assumptions, respectively (also discussed by Smith et al., 2009; Hinchliff and Smith, 2014b).

As phylogenomics has developed over the last decade, alternative methods have been introduced to accommodate the increased amounts of data and inherent gene-tree–species-tree conflicts. These methods measure the concordance of gene trees (broadly referring to a phylogeny from any subsampled genomic region), including the internode certainty (IC) and tree certainty (TC) scores (Rokas et al., 2003; Salichos et al., 2014; Kobert et al., 2016; Zhou et al., 2017), Bayesian concordance factors (Ane et al., 2006), and other concordance measures (Allman et al., 2017). These scores were developed around the central concept of a branch support statistic that measures concordance of various trees with a particular tree hypothesis. This perspective offers much for partitioning phylogenetic discordance and analyzing larger alignments more rapidly in a phylogenomic coalescent-based framework. Unfortunately, though relevant to genomic data sets, they may not be as suitable for large-sparse alignments.

Finally, quartet methods—in particular, quartet puzzling methods—have been developed for phylogenetic reconstruction (Strimmer et al., 1997; Strimmer and von Haeseler, 1997; Ranwez and Gascuel, 2001; Allman and Rhodes, 2004; Chifman and Kubatko, 2014; Mirarab et al., 2014; Zwickl et al., 2014) and support (e.g., “reliability values”; Strimmer et al., 1997; Strimmer and von Haeseler, 1997). More recently, quartet procedures have been explored to facilitate sampling of large-sparse alignments (Misof et al., 2013) and as part of coalescent-based quartet inference methods (Stenz et al., 2015; Gaither and Kubatko, 2016; Sayyari and Mirarab, 2016). These quartet methods benefit from the speed advantages of a smaller alignments and the statistical consistency of quartet trees, which avoid complex lineage sorting issues that occur with more speciose phylogenies (Rosenberg, 2002; Degnan and Salter, 2005).

Despite the wide array of approaches to branch support quantification briefly discussed above, few measures (excepting concordance methods) accommodate multiple histories and distinguish different causes of poor support for a branch in the phylogeny (e.g., multiple supported-but-conflicting phylogenetic relationships versus low information). Being able to identify a branch as having a strong consensus and a strongly supported secondary evolutionary history would provide valuable insight into the green plant tree of life (among many other groups; see also Brown and Lemmon, 2007).

Here, we describe the Quartet Sampling (QS) method (summarized in Fig. 1 and Table 1), which blends aspects of many of the methods described above and leverages the efficiency of quartet-based evaluation. The goal of the QS method is to dissect phylogenetic discordance and distinguish among lack of support due to (1) low information (as in NBS and PP), (2) discordance as a result of lineage sorting or introgression (as in concordance measures), and (3) misplaced or erroneous taxa (a.k.a. “rogue taxa”; Wilkinson, 1996; Aberer et al., 2012). In many modern phylogenetic and particularly phylogenomic studies, these causes of discordance are frequently surveyed and reported separately (e.g., Xi et al., 2014a; Wickett et al., 2014; Yang et al., 2015; Pease et al., 2016b; Walker et al., 2017c). Quartet Sampling provides a unified method for their execution, interpretation, and reporting. Additionally, the QS method offers a viable means to describe branch support in large phylogenies built from sparse alignments (10,000–30,000 tips with >80% missing data), which are generally intractable for Bayesian analysis.

In this study, we (1) describe the features, parameters, and interpretation of the QS method, (2) validate the QS method with simulations, and (3) apply the QS method to recently published large-sparse and phylogenomic data sets at timescales spanning from *Viridiplantae* (green plants) to subgeneric clades. We demonstrate that the QS method is a flexible and computationally tractable method for examining conflict and support in large data sets. While not a panacea, we argue that the QS framework makes important steps in addressing many of the issues of branch support discussed above and hope it encourages additional discussion, testing, and innovation of new phylogenetic evaluation methods. The results presented herein contribute to the broader discussion about moving the plant tree of life beyond the goal of resolving a single, universal species tree (Hahn and Nakhleh, 2015; Smith et al., 2015) and into a future in which we more fully explore and appreciate the complex “multiverse” of evolutionary histories manifest throughout the plant tree of life.

## **<h1>MATERIALS AND METHODS**

### **<h2>Quartet Sampling—**

The Quartet Sampling (QS) procedure outlined here was inspired by aspects from several quartet-based and concordance methods, most particularly the process originally outlined by Hinchliff and Smith (2014b). The QS method takes an existing phylogenetic topology (which can be inferred by any method) and a molecular data set (not necessarily the one that generated the

phylogeny) and separately evaluates one or more internal branches on the given phylogeny. The QS method (Fig. 1) was designed to rapidly and simultaneously assess the confidence, consistency, and informativeness of internal tree relationships, and the reliability of each terminal branch.

For a given phylogeny, each observed internal tree branch partitions the tree into four nonoverlapping subsets of taxa (Fig. 1A). These four sets of taxa (called a meta-quartet by Zhou et al., 2017) can exist in three possible relationships: the concordant relationship that matches the configuration in the given topology, and two alternative discordant configurations. The QS method repeatedly and randomly samples one taxon from each of the four subsets and then evaluates the likelihood all three possible phylogenies given the sequence data (either the full alignment or a randomly sampled gene partition) for the randomly selected quartet spanning that particular branch.

For each quartet sampled at a focal branch, the likelihood is evaluated (using the aligned sequence data) for all three possible topologies that these four sampled taxa can take (currently using RAxML or PAUP\*, though other likelihood calculators could be substituted; Swofford and Sullivan, 2003; Stamatakis, 2014). The quartet topology with the best likelihood is then recorded and tabulated across all replicates. This process generates a set of counts (across all replicates per branch) where either the concordant or each of the two discordant relationships had the best likelihood. This procedure can be carried out by evaluating the likelihood of the complete alignment for each quartet (i.e., in a single-matrix framework) or by randomly sampling from individual gene/partition alignments from a multigene or genomewide alignment (i.e., in a multigene tree coalescent framework).

Several refined options can be specified. For example, a minimum number of overlapping non-empty sites for all four taxa involved in a quartet can be specified to ensure calculations are performed on data rich subsets. Additionally, a parameter of a minimum likelihood differential may be set. If the most-likely topology (of the three) does not exceed the likelihood of the second-most-likely phylogeny by the set threshold, then the quartet is considered “uninformative” and tabulated separately. In summary, the QS method generates counts of the three possible topologies (and uninformative replicates) for each internal branch by sampling replicates using unique quartets of taxa spanning the particular branch.

The QS method uses these resampled quartet tree counts to calculate three scores for each internal branch of the focal tree (Fig. 1B, Table 1; Appendix S1, see online Supplemental Data with this article). The Quartet Concordance (QC) score is an entropy-like measure (similar to the ICA score; Salichos et al. 2014) that quantifies the relative support among the three possible resolutions of four taxa. When the most commonly sampled topology is concordant with the input tree, then the QC value is positive in the range (0,1]. Thus, QC equals 1 when all quartet trees are concordant with the focal branch. When one of the discordant topologies is the most commonly resampled quartet, QC value is negative in the range [-1,0), approaching -1 when all quartet trees are one of the two discordant phylogenies. The QC equals 0 when support is evenly split among the three alternative topologies (or evenly split between two if only two of the three possible are registered as having an optimal likelihood across all replicates).

The Quartet Differential (QD) score uses the logic of the  $f$  and  $D$  statistics for introgression (Reich et al., 2009; Green et al., 2010; Durand et al., 2011; Pease and Hahn, 2015) and measures the disparity between the sampled proportions of the two discordant topologies (though with quartet topology proportions, rather than site frequencies). The QD score does not specifically quantify introgression nor identify introgressing taxa, but does indicate that one alternative relationship is sampled more often than the other is. Low values of QD indicate that there is one preferred topology between the two discordant topologies, a potential indication on the given branch of a biased biological process beyond background lineage sorting, including confounding variables such as introgression, strong rate heterogeneity, heterogeneous base compositions, etc. QD varies in the range [0,1] with a value of 1 meaning no skew in the proportions of the two discordant trees and the extreme value of 0 meaning that all discordant trees sampled are only from one of the two possible alternative relationships.

The Quartet Informativeness (QI) score quantifies for a given branch the proportion of replicates where the best-likelihood quartet tree has a likelihood value that exceeds the quartet tree with second-best likelihood value by a given differential cutoff. This score clarifies that replicates are not counted as being concordant or discordant when the molecular data are effectively equivocal on the topology (i.e., when two or all three of the three possible quartet topologies have nearly indistinguishable likelihood scores). Quartet Informativeness is measured in the range [0,1], which indicates the proportion of sampled quartets that exceeded the cutoff. A QI value of 1 means all quartets are informative, while a value of 0 indicates all quartets were



uncertain (i.e., no significant information for the given branch). The QI measure of branch informativeness works in conjunction with QC and QD to distinguish between branches that have low information versus those with conflicting information (i.e., high discordance).

Finally, for each terminal taxon, a Quartet Fidelity (QF) score is calculated to report the proportion of total replicates (across all branches tested) including the given taxon that resulted in a concordant quartet topology. The use of QF is therefore similar in approach to a “rogue taxon” test (Wilkinson, 1996; Aberer et al., 2012). However, an important distinction is that uses taxonomically complete bootstrap replicates to compute these scores rather than resampled subtrees, and thus are subject to the same issues as bootstrap scores themselves in phylogenomic analyses (i.e., the RogueNaRok algorithm will not report rogue taxa when all bootstrap scores are 100). For a given taxon, the QF score is measured in the range [0,1] as the proportion of quartet topologies involving the taxon that are concordant with the focal tree branch. Therefore, a QF value of 1 indicates a given taxon always produces concordant topologies across all internal branches where it was sampled for in a quartet. QF values approaching zero indicate mostly discordant topologies involving this taxon and may indicate poor sequence quality or identity, a lineage-specific process that is distorting the phylogeny, or that the taxon is significantly misplaced in the given tree. Note that QF differs specifically from QC, QD, and QI by being a taxon-specific test across internal branch tests rather than an internal branch-specific test.

Collectively, these four tests provide the means to distinguish the consistency of information (QC), the presence of secondary evolutionary histories (QD), the amount of information (QI), and the reliability of individual taxa in the tree (QF; Fig. 1B and see Table 1). The QS tests disentangle these effects rather than have them conflated under a single summary score as in standard measures of phylogenetic support. A full technical description of the QS method is included in Appendix S1.

## **<h2>Implementation of QS—**

We implemented the above procedure in a Python-based program called quartetsampling (Pease et al., 2017), which samples an alignment randomly to generate many representative quartet topology replicates for each internal branch in a corresponding focal tree. This procedure has a number of advantages over NBS for larger data sets. First, unlike NBS and RBS, alignment columns are not resampled, which allows sparse alignments to be used. Second, the number of

likelihood calculations that are required is the number of internal branches in the tree multiplied by the number of replicates per branch multiplied by three possible topologies. Therefore, computation time scales linearly with the number of taxa. Since individual replicates are independent, the computations can be readily parallelized across processors and systems (with results combined later), allowing QS to be efficiently applied to large alignments beyond the practical limits of NBS and PP. The most extensive computational time was for the 31,749-taxon data set of Zanne et al. (2014b) (see below), which we ran on the Wake Forest University DEAC high-performance cluster using 8 nodes with 16 CPU each. This analysis completed 200 replicates for the full tree in 13 h. Smaller genome-wide data sets finished 1000 gene-tree replicates on four-core desktops in approximately 12 h. The conventional multigene data sets took only a few minutes to a few hours to run on a standard desktop.

Although the Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-aLRT; Guindon et al., 2010) was by far the fastest method we consider here, the QS was fast enough for large-scale analyses. QS can also be applied separately to individual focal branches, allowing a more thorough exploration of particular branches of interest. Furthermore, QS does not require the tree tested to be the maximum likelihood topology, a requirement for SH-aLRT. For our simulated data, we found that performing 200 QS replicates per branch was adequate to achieve a low variance in the QS score (Fig. 2A). As would be expected, more replicates per branch should generally be used for larger trees to sample a greater fraction of the total possible quartets. Some branches, especially in large trees, may be entirely unsupported by the alignment due to a lack of alignment overlap among any possible combination taxa from the four subsets (Fig. 1A). Therefore, no phylogenetic information exists to inform the branch (i.e., they are not “decisive” sensu Steel and Sanderson, 2010). The QS procedure identifies these branches, rather than discarding them or ambiguously labeling them as having “low support”.

## ***<h2>Guidelines for interpretation of QS support values—***

An important consideration with any measure used to ascertain confidence is the precise interpretation. We provide a concise visual description of the tests (Fig. 1) and a table describing example scores and their interpretations (Table 1). Particularly notable is that the QS method not only can “support” or “fail to support” a given branch hypothesis, but also can offer “counter-support” for an alternative branch (as in the IC/ICA scores; Salichos et al., 2014; Kobert et al.,

2016; Zhou et al., 2017). Therefore, even “inaccurate” branch hypotheses can offer information as “counter-support” for an alternative quartet topology (i.e., the degree of negativity of the QC score; see Table 1).

The QS scores we have described calculate the sensitivity of the resolution of a particular branch to different combinations of taxa sampled around that branch. Each QS replicate calculates whether the four sampled taxa support the resolution of the branch found in the tree over the alternative resolutions. This framework is similar to interpretations of taxon jackknife analyses for outgroup sensitivity (e.g., Edwards et al., 2005) and the IC score when used with incomplete trees (Kobert et al., 2016; Zhou et al., 2017). We argue that this interpretation is richer in information than NBS. In simulations, the QC score also appears to more conservatively and accurately assign high support values to branches that are present in the true tree (i.e., relatively low false positive rates, at least when the likelihood threshold is small in the range of  $\sim 2$  used here; Appendix S2). QC scores are particularly helpful for clarifying strength of support for branches with concordant tree frequencies not close to 1 (Appendix S3).

## **<h2>Generation and evaluation of simulated phylogenies—**

We first tested the QS method by generating simulated phylogenies under the pure birth (birth = 1) model of evolution with 50, 100, and 500 tips using `pxbdsim` from the `phyx` toolkit (Brown et al., 2017). Using these trees, we generated 1000-bp alignments (no indels) under the Jukes–Cantor model with `INDELible` v. 1.03 (Fletcher and Yang, 2009). Trees were scaled so that the average branch lengths were about 0.2, based on the observation that this generated reasonable trees with most branches recovered correctly from ML analyses. Using the same procedure, we also simulated trees with 500 tips and associated alignments with 10 nucleotide partitions, each with 500 sites under the Jukes–Cantor model. We simulated both the full alignment with partitions and a modified randomly resampled sparse alignment to examine the behavior of QS in the presence of missing data (see Appendix S1 for details). These partitioned and sparse alignments had the same qualitative features as the full alignment.

Unlike the NBS method, which generates a set of trees from which branch support is estimated, the QS method requires only a single input topology for which branch support will be measured. We calculated QC, QD, QI, and QF scores for the true underlying tree as well as the ML tree generated by `RAxML`, but we focus on results for the ML tree. To examine how the

number of replicates affects the QS precision, we conducted simulations where we increased the number of replicates for randomly drawn branches in the simulated trees (Fig. 2A; Appendix S4). Based on these simulations, we elected to use 200 replicates per branch, since the variance in the QC score was generally low across all tree sizes when this many replicates were performed. We used RAxML and PAUP\* to estimate the ML for the three alternative topologies for each QS replicate (using the “-f N” option and the GTR+ $\Gamma$  model in RAxML). We also calculated branch-specific QC/QD/QI and taxon-specific QF scores using likelihood differential cutoffs of  $\Delta L = 0$  (no filtering) and  $\Delta L = 2.0$ , which requires stronger conflicting signal to interpret branches in the input tree as unsupported.

Additionally, we generated a simulated 20-taxon tree with variable branch lengths using the tool pxbdsim from the phyx toolkit (Brown et al., 2017) (Appendix S5). For 100 replicates, we generated twenty 5-kb nucleotide sequences over this tree using the program ms (Hudson, 2002), inferred a concatenated tree using RAxML (Stamatakis, 2014), and used this inferred tree and simulated alignment as the inputs for QS. Population parameters were set at  $\mu = 1 \times 10^{-8}$  and  $N_e = 10^5$ . To simulate increasing amounts of ILS, we shortened the times between speciation events by scaling all branch lengths by factors ranging from 0.5 to 10 and repeated these simulations (Fig. 2C). Additionally, using the original tree scaled by a factor of 2, we added introgression of varying intensity between “taxon\_6” and “taxon\_7” (using the migration parameter in ms from 0 to  $1.4/4N_e$  migrants per generation). Additional details can be found in Appendix S1.

## <h2>Testing of empirical data sets—

We evaluated several recent large-scale phylogenies: a 103-transcriptome data set spanning *Viridiplantae* from Wickett et al. (2014; hereafter “WI2014”), two large-sparse phylogenies spanning land plants from Hinchliff and Smith (2014b; “HS2014”) and Zanne et al. (2014b; “ZN2014”), and phylogenies spanning Magnoliophyta (angiosperms) with hundreds of genes from Xi et al. (2014a, “XI2014”) and Cannon et al. (2015b, “CN2015”). Additionally, to demonstrate the utility of this method at medium and short time scales, we evaluated two whole transcriptome data sets from the wild tomato clade *Solanum* sect. *Lycopersicon* from Pease et al. (2016b, “PE2016”) and carnivorous plants from the order Caryophyllales from Walker et al. (2017c, “WA2017”). Finally, we tested this method on a more conventionally sized multilocus

data set from Polypodopsida (ferns) from Pryer et al. (2016b, “PR2016”), such as might appear in many phylogenetic studies of large subgroups. Data for these studies were obtained from the Dryad (<http://www.datadryad.org>) and CyVerse (<http://www.cyverse.org>) databases (Hinchliff and Smith, 2014a; Matasci et al., 2014; Xi et al., 2014b; Zanne et al., 2014a; Cannon et al., 2015a; Pease et al., 2016a; Pryer et al., 2016a; Walker et al., 2017b; additional details and results in Appendix S1).

We performed QS using 200 individual gene trees for XI2014 and WA2017, and 1000 gene trees for PE2016 and WI2014. In this gene tree mode of QS, quartets are sampled as usual, but only the likelihoods of individual gene sequence alignments are compared. These phylogenies were all evaluated using a minimum alignment overlap per quartet of 100 bp and a minimum likelihood differential of 2 (i.e., the optimal tree’s log-likelihood must exceed the secondmost likely tree by a value of at least 2). We also calculated the phylogenies with and without partitioning in RAxML, but in all cases the partitioned data sets did not qualitatively differ from the results of the unpartitioned data sets.

We also either recalculated other measures of branch support or used values from the published studies for comparison to the QS method for each phylogeny, except HS2014 and ZN2014 where the size and sparseness of the data sets prohibited the calculation of other measures of support. For the data sets from CN2015, PR2016, WA2017, and XI2014, 100 replicates each of RAxML NBS and SH-test were performed. Additionally, PP scores for PR2016 were calculated using MrBayes (Ronquist et al., 2012), and IC scores for WA2017 were calculated using RAxML. For PE2016 and WI2014, RAxML NBS, MP-EST, or IC scores were taken from published values. Finally, we also calculated QF scores and rogue taxon scores using RogueNaRok (Aberer et al., 2012) to compare these two measures, particular for the large-sparse ZN2014 data set (for details and results, see Appendix S1). All data used and results from both simulations and empirical data sets are available in the Dryad Digital Repository (<http://dx.doi.org/10.5061/dryad.6m20j>).

## **RESULTS AND DISCUSSION**

### ***Simulation analyses—***

We tested the consistency and reliability of QS on a set of simulated phylogenies. The QC scores converged (with decreasing variance as expected) on a consistent mean value for each branch as

the number of replicates increased (Fig. 2A). Sampling 200 quartets per branch reduced the variance to less than  $3 \times 10^{-3}$  in all cases, and can be seen as a generally a reasonable number of replicates. As these are branch-specific tests, branches of interest can be tested individually at much higher numbers of replicates without the need to retest the entire tree.

We also simulated sequences over a standard phylogeny (Appendix S5), and then simulated increasing ILS by shortening branch lengths and introgression via migration. As expected, QC scores that measured concordance decreased in both cases due to the increased presence of discordant sites and QD scores that measure skew in discordance decreased dramatically with increasing directional introgression (Fig. 2B). We also found that while QC and QD both measure discordance levels, they are not strictly correlated with each other or the raw frequency of the concordant quartet (Appendices S2, S6). As QC goes to the limits of its range  $[-1,1]$ , QD values tend to have more extreme values due to a lack of discordant trees (QC near 1) or high frequency of one discordant tree (QC near  $-1$ ). Applying a minimum log-likelihood differential threshold to small trees tended to push scores toward extremes, resulting in more 0s and 1s (Appendix S2). Finally, we found that those data sets with lower QF score generally identified more rogue taxa than inferred by RogueNaRok, despite the different data inputs and analysis frameworks (Appendix S1).

## <h2>QS analyses of major land plant lineages—

The primary goal of this study was to use QS to reanalyze and compare several recent speciose and phylogenomic data sets to address ongoing debates of phylogenetic relationships in the green plant tree of life. We used QS methods to evaluate two of the most speciose phylogenies of land plants currently available (Hinchliff and Smith [2014b; Fig. 3]; Zanne et al. [2014b; Fig. 4]), and one of the most comprehensive phylogenies of *Viridiplantae* (Wickett et al., 2014; Fig. 5). QS analyses were able to provide a broad scale summary of the stability of the data sets.

As expected, given the sparsity of the matrices for HS2014 and ZN2014 (96% and 82% missing characters, respectively), the proportion of informative quartets was low in both cases (mean QI of 0.15 and 0.35, respectively). Overall, the mean QC for the HS2014 (0.15; interquartile range (IQR) =  $[-0.13, 0.46]$ ) and ZN2014 (0.17; IQR =  $[-0.10, 0.63]$ ) were low compared to the less speciose phylogenies (Fig. 2C; Appendix S7). Notably, we found 33.4% and 29.8% of branches in HS2014 and ZN2014, respectively, had QC values less than  $-0.05$ . Thus,

consensus phylogenies reported not just “low support” for a third of the branches, but went further to report “counter-support” (i.e., a negative QC score) for one of the two alternative topological arrangements. Most major plant taxonomic groups showed strong support in HS2014 and ZN2014, and all major groups showed strong support in WI2014 (Table 2). In contrast to strong support for major groups themselves, we found low support along the “backbone” relating these groups, in a manner consistent with most previous phylogenies of land plants.

The relationships among Marchantiophyta, Bryophyta (mosses), Anthocerotophyta, lycophytes, and euphyllophytes (i.e., ferns and seed-bearing plants) have been matters of ongoing debate (Shaw et al., 2011). HS2014 inferred Bryophyta as a lineage separate from the remaining land plants, but showed counter-support (negative QC =  $-0.04$ ) for a branch defining a common ancestor of liverworts with all other land plants to the exclusion of mosses (Fig. 3B). This result suggested that the most common quartet branch among the replicates was not the branch displayed in the published tree. By contrast WI2014 showed strong support (with a high QC =  $0.67$ ) for a common ancestor of mosses and liverworts (Fig. 5). ZN2014 indicated weak support (low positive QC =  $0.15$ ) for the branch separating mosses and liverworts from the rest of land plants. Therefore, while the topology of HS2014 was consistent with many previous phylogenies (Nickrent et al., 2000; Qiu et al., 2006; Chang and Graham, 2011), the QS results collectively supported the alternative configuration of mosses and liverworts as sister groups (Fig. 6A; see also Renzaglia et al., 2000; Zhong et al., 2013a).

In all three data sets, the monophyly of vascular plants (tracheophytes) was strongly maintained, even with the inclusion of *Selaginella* with its unusual GC content (Banks et al., 2011). The branch leading to *Selaginella* often had a lower QD value, possibly because of this biased composition, but a higher QF value, suggesting that it was not a misplaced (“rogue”) taxon. We also observed substantial discordance and counter-support for relationships tested among various bryophyte groups and key taxa in HS2014, possibly indicative of substantially under-appreciated hybridization among mosses (Nylander et al., 2004).

## <h2>QS analyses of ferns—

The branch establishing a euphyllophyte common ancestor of ferns and seed-bearing plants showed low QC scores and high QD scores in both HS2014 and ZN2014, indicating a weak consensus but with little indication of an alternative history (Table 2). Within ferns, the

arrangement of major clades in ZN2014 (Fig. 4E) was mostly consistent with the recently published phylogeny by the Pteridophyte Phylogeny Group (PPG I, 2016). Those clades whose relationships were counter-supported (Marratiales, Salviniiales, Hymenophyllales) were discordant with the PPG-I consensus and other recent phylogenies (Pryer et al., 2004; Testo and Sundue, 2016), which demonstrated the diagnostic utility of QS in highlighting suspect relationships. Some key areas of known high uncertainty (e.g., *Saccoloma*, *Lindsaea*, and *Equisetum*) were also highlighted with low or negative QC scores.

While QS was designed for large data sets, we also found that QS performed well on smaller multigene data sets conventionally used for systematics studies. The QS scores for PR2016, with a 5778-bp alignment, were more conservative, but confirmed the conclusions of Pryer et al. (2016b) regarding the monophyly of maidenhair ferns (*Adiantum*) and its placement in a clade with the vittarioids (Fig. 7). This analysis also revealed some counter-supported nodes (negative QC values) within the genus *Adiantum*.

## <h2>QS analyses of gymnosperms—

Another question that has attracted substantial historical debate is the relationships among the major extant gymnosperm lineages and angiosperms. Under QS evaluation, all four testable data sets indicated strong support for monophyly of gymnosperms (Table 2), but with variable relationships among the major lineages within gymnosperms. ZN2014 and WI2014 inferred a common ancestor of *Ginkgo* and cycads (consistent with Bowe et al., 2000; Qiu et al., 2006; Lee et al., 2011; Xi et al., 2013). While the HS2014 topology places cycads as sister to the remaining gymnosperms (i.e., not monophyletic with *Ginkgo*), the QS evaluation counter-supports this relationship. Therefore, even though HS2014 and WI2014 differed from ZN2014 in the topological relationship of these taxa, the QS analyses of these data sets indicated a consistent message of a *Ginkgo* and cycads common ancestor separate from the rest of gymnosperms (Fig. 6B).

This pattern of disagreeing topologies but consistent QS interpretation was observed again in the placement of Gnetales relative to the conifer lineages (Fig. 6C). ZN2014 showed a common ancestor of Gnetales and Pinales (consistent with Lee et al., 2011). Even though a conflicting Gnetales and Pinaceae ancestor (distinct from other conifers) appeared in both HS2014 and WI2014 (i.e., the “Gnepine” hypothesis; Bowe et al., 2000; Xi et al., 2013), the negative-QC/low-



QD scores in both cases indicated counter-support for a “Gnepine” ancestor and a strongly support alternative history (QC/QD =  $-0.19/0.56$  and  $-0.67/0.0$ , respectively). Collectively, these results suggest the monophyly of Pinales, but also offer some (albeit weak) evidence that warrants further examination of possible gene flow between Gnetales and Pinales.

## **<h2>QS analyses of ANA-grade angiosperms—**

Few issues in angiosperm evolution have garnered more recent debate than the relationship among the ANA-grade angiosperms (Qiu et al., 1999), which include *Amborella*, Nymphaeales, and Austrobaileyales. Two questions surround the evolutionary history of the ANA-grade angiosperms. First, what are the relationships among these lineages? Second, are the longstanding disagreements in inference of these relationships the result of genuine biological conflict (i.e., introgression, horizontal transfer, etc.), limitations in the data, or a methodological artifact (i.e., due to the depth of this branch, the monotypic status of *Amborella*, and/or the rapidity of the angiosperm radiation)?

On the first question, QS analyses of the data sets here lack support for “Nymphaeales-first” but finds support for both *Amborella*+Nymphaeales and “*Amborella*-first” (as found also by The Amborella Genome Project, 2013). While the resolutions of consensus phylogenies differ between the four testable data sets (WI2014 with “*Amborella*-first” hypothesis, ZN2014 with “Nymphaeales-first”, and HS2014 and XI2014 with *Amborella*+Nymphaeales), the branches surrounding the ANA grade were all counter-supported (QC < 0) and biased in their discordance (QD < 0.2; Fig. 6D). ZN2014 offers weak support for *Amborella*+Nymphaeales, while XI2014 counter-supports this relationship. If this first question is to be resolved, our results indicate additional data sets or different analyses will be required.

On the second question, our analyses show low QD values that suggest a conflicting phylogenetic history may be present. Other studies have found bryophyte mitochondrial sequences present in *Amborella* (Rice et al., 2013; Taylor et al., 2015), which establishes the potential for introgression in these lineages. Overall, (1) the intense efforts to address these relationships without a resulting broad community consensus, (2) evidence of long-range introgression, and (3) the QS results presented here together suggest that a greater understanding of the evolution of ANA-grade angiosperms likely lies in an examination of a more complex

evolutionary history at this key transition, instead of in continuation of the debate over appropriate sampling or models (see also discussion by Shen et al., 2017).

## <h2>*QS analyses of “core angiosperms”*—

The three “core angiosperm” lineages (eudicots, monocots, and magnoliids) have transformed the biosphere, and thus a better understanding of the timing and order of their origins is of key concern. Consensus topologies disagree between ZN2014, WI2014, and XI2014 (with magnoliid+eudicot clade; Figs. 4B, 5, 6E; Appendix S8) and HS2014 (with eudicot+monocots; Fig. 3B). However, the QS analyses of HS2014 showed counter-support of an exclusive common ancestor of eudicots and monocots, suggesting that, despite disagreement among topologies, QS scores support a common ancestor for magnoliids and eudicots to the exclusion of monocots. Additionally, the placement of Chloranthaceae seems inextricably linked with the relationships of the three core angiosperm groups (see discussion by Eklund et al., 2004). However, the placement of this family remains unresolved by QS, since all tested configurations showed negative QC-value counter-support (Table 2).

## <h2>*QS analyses of monocots*—

In general, the arrangement of monocot orders in both HS2014 (Fig. 3C) and ZN2014 (Fig. 4C) agreed with recent consensus phylogenies (Givnish et al., 2010, 2016; Barrett et al., 2015; McKain et al., 2016). Two exceptions are the placement of Liliales (Table 2), and general inconsistency of commelinid orders. From the QS results, we would cautiously infer that (1) the relationships among the commelinids are still unknown, (2) there may be uncharacterized secondary evolutionary history distorting the phylogenetic placement of these groups, and (3) the variable data from both Liliales and Arecales may be creating a joint effect in causing inconsistency in the phylogenetic inference.

In Poaceae, QS analyses highlight the well-characterized discordance and complex relationships (e.g., Washburn et al., 2015; McKain et al., 2016). Even if someone were completely unfamiliar with the known controversies in monocots, QS scores would make abundantly clear this area of the phylogeny had highly conflicted data. The “BOP” clade itself and many clades within the “PACMAD” clade were counter-supported by negative QC values in HS2014 and ZN2014. However, low QI values were observed in both HS2014 and ZN2014 for this clade,

indicating that both data sets contain poor information. Therefore, QS serves as an effective diagnostic tool for identifying conflicted portions of larger phylogenies.

## <h2>*QS analyses of non-rosid/asterid eudicots—*

QS analyses are capable of identifying conflict and discordance due to rapid radiations as demonstrated well for the relationships among the superasterid groups (Caryophyllales, Berberidopsidales, Santalales, and asterids). A common pattern was found in HS2014, WI2014, XI2014, and ZN2014 of near-zero QC values ( $-0.03$  to  $0.08$ ) that indicate weak consensus for the given relationships, strong QD values ( $0.97$ – $1$ ) that indicate no strongly competing alternative history, and low QI values ( $0.14$ – $0.51$ ) that indicate low information for branches. This result led to a consensus QS interpretation of poor phylogenetic information, likely due to the rapid radiation of these lineages. Generally, these phylogenies tended to support weakly the controversial placement of Caryophyllales as most closely related to the eudicot ancestor.

## <h2>*QS analyses of rosids and asterids—*

Analysis of the rosids confirms that the QS method is capable of identifying rogue taxa. The QS scores identified a poorly supported relationship in HS2014 between *Cynomorium* and Cucurbitales ( $QC = -0.31$ ). *Cynomorium*, a nonphotosynthetic parasitic plant with unusual morphology, has been placed tenuously and variably in groups as diverse as Rosales (Zhang et al., 2009) and Saxifragales (Nickrent et al., 2005). This “rogue” status was corroborated by a below-average QF score of  $QF = 0.18$  (mean  $0.21$  for HS2014). This low QF score means that for quartets that include *Cynomorium* as a randomly sampled taxon, only 18% produced a quartet topology concordant with the HS2014 tree.

Published phylogenies of asterids have indicated disagreement and substantial discordance (Soltis et al., 2011; Beaulieu et al., 2013; Refulio-Rodriguez and Olmstead, 2014). QS scores from ZN2014 supported the unusual hypothesis of an exclusive common ancestor of Ericales and Cornales, weakly supported the campanulid clade, and counter-supported a common lamiid ancestor. The arrangement of families within Asterales either roughly conformed to Soltis et al. (2011) and Beaulieu et al. (2013), or counter-supported branches ( $QC < 0$ ) that did not agree with these consensus phylogenies (see also analysis of CN2014; Appendix S9). However, most of the branches that define relationships among asterid orders in ZN2014 were counter-supported by the

data, though most had QC and QD values close to zero. This indicated a scenario of a rapid radiation rather than hybridization (though these are not mutually exclusive).

## <h2>QS analyses of shallow-timescale phylotranscriptomic datasets—

So far, we have demonstrated the utility of Quartet Sampling on large, sparse, and conventional multigene alignments, which are often computationally intractable with other support measures. We have also shown for WI2014 that a relatively large and fully occupied matrix from deep-timescale transcriptomic data can also be evaluated by QS. However, the QS method also can rapidly evaluate phylogenetic support on genome-wide data sets with little missing data for shorter evolutionary timescales. We tested the QS method on two phylotranscriptomic datasets for the wild and domesticated tomato clade *Solanum* sect. *Lycopersicon* (Fig. 8A; Pease et al., 2016b) and carnivorous plants spanning the Caryophyllales (Fig. 8B; Walker et al., 2017c).

The *Solanum* phylogeny of Pease et al. (2016b) was inferred from the alignment of 33,105,168 nucleotide sites for 30 populations spanning all 13 wild and domesticated tomato species, and two outgroup species. As described by Pease et al. (2016b), this data set contains a high level of phylogenetic discordance, but had a consensus phylogeny with 100% NBS support at all but two branches. However, gene tree analysis of this group showed evidence of massive phylogenetic discordance. When we applied QS to this phylogeny using the entire alignment, scores for many branches were also perfect (i.e., 1/–/1; Table 1). However, several of the other branches in the “Peruvianum group” species complex had lower QS scores in the full alignment (Fig. 8A). When gene trees were used (a random gene and quartet of taxa were chosen for 1000 QS replicates), all branches had  $QC < 1$  in a manner consistent with the gene tree discordance found previously in this clade. We also observed the presence of low QD values within the major subgroups reported for this clade, indicating the presence of introgressive gene flow. In contrast, nodes defining the major subgroups showed high QC and QD values that indicated strong monophyly. These scores accurately captured the low discordance between groups versus high discordance within the major groups found by Pease et al. (2016b).

Most notably, the tree shown in Fig. 8A includes *S. huaylasense* 1360. This population has been known (from both Pease et al. [2016b] and other data sets) mostly likely to be a hybrid between populations from the green-fruited and red-fruited lineages (essentially those accessions above and below LA1360, respectively, in Fig. 8A). Thus, the inclusion of this putative hybrid

lineage distorted the phylogeny as tree inference methods tried to cope with inherited and introgressed alleles from two separate groups to place this accession in a consensus location on the tree. While NBS scores were high for the branches surrounding the placement of LA1360, QS showed negative QC scores and low QD scores (QD = 0 for full alignment). The low QD supports the presence of the alternative phylogenetic history that has been previously corroborated by other studies and the negative QC indicates counter-support for the placement of this accession (see additional discussion in the Supplementary Results of Pease et al., 2016b). These data show that QS was able to distinguish between consistently supported relationships and branches known to have conflict due to introgression (whereas NBS does not).

An analysis of transcriptomes of carnivorous plants from Caryophyllales (Fig. 8B; Walker et al., 2017c) was also enhanced by QS. The near-zero QC scores and low QD (0.32) scores for the ancestor of a clade containing *Plumbago* and *Nepenthes* for gene trees supported the hypothesis of Walker et al. (2017c) that introgressive gene flow may have occurred among these lineages. Evidence for placing *Drosophyllum* among the carnivorous Caryophyllales has been previously tenuous, and the QS analysis showed not only a low QF value (0.76) compared to the WA2017 mean QF (0.89) for this taxon, but also low-QC/low-QD values for the two branches that form the clade with *Ancistrocladus* and *Nepenthes*. As with *Solanum*, this example demonstrates how QS scores highlighted an entire sector of the topology that may be distorted by the inclusion of a taxon with a likely secondary evolutionary history (i.e., possible introgression).

## **<h2>Limitations and directions forward—**

Quartet Sampling was designed to evaluate phylogenetic information efficiently and to highlight conflict for one or more branches in a phylogeny. In the presentation here, QS is used to evaluate a single topology, and not for comparing alternatives topologies or performing any optimizations that might maximize QS scores. Therefore, QS does not suggest topological rearrangements and is purely evaluative. These and other directions should be explored in future studies as researchers develop more ways to examine uncertainty in large data sets.

Concurrently with our study, Zhou et al. (2017) proposed the Q-IC method, an approach similar to QS. Both approaches use quartets to evaluate a focal tree. Both approaches can be used in a single-matrix or multigene tree framework, implemented in Q-IC by sampling from either a single tree distribution or from a gene tree set, and implemented in QS by analyzing either the

whole alignment or by also randomly sampling individual gene-quartet combinations (as shown in Fig. 8). One key difference is that QS evaluates the relative likelihood of all three possible quartet configurations for each branch based on the alignment data set, while Q-IC evaluates only the quartet topologies sampled from a data set of topologies from “evaluation trees” (i.e., individual gene trees or a bootstrap/posterior distribution). These differences in data evaluation might make these approaches sensitive to different error types (e.g., gene tree topological estimation error versus likelihood estimation errors). Overall, we find QS and Q-IC complementary in their approaches and their appropriateness dependent upon the data available and types of questions being asked.

## <h1>CONCLUSIONS

We reanalyzed several long-contested, key conflicts in the green plant tree of life and describe a framework for distinguishing several causes of low phylogenetic branch support. For large data sets, traditional measures such as the bootstrap or posterior probabilities may be computationally intractable, exhibit irregular behavior, or report high confidence despite substantial conflict. The QS framework provides a tractable means to analyze sparse data sets with tens of thousands of taxa but poor sequence overlap. QS provides a key function that has been missing from other support measures, namely the ability to distinguish among different causes of low support that commonly occur in modern molecular phylogenies. We demonstrate this by reporting the existence of multiple conflicting but supported evolutionary histories at several key points in the green plant tree of life (e.g., the placement of *Amborella*, possible widespread gene flow in the monocots, and notoriously difficult-to-place taxa like *Cynomorium*). We hope that our discussions here will also lead to the development of other means for parsing the information contained within exponentially expanding molecular data sets. The artist Man Ray once remarked: “We have never attained the infinite variety and contradictions that exist in nature.” Overall, the picture painted by QS is one of substantial contradiction, but this conflict can be a richly informative (not just confounding) illustration of the interwoven evolutionary histories contained within the tree of life.

## <h1>ACKNOWLEDGEMENTS

The authors thank Ya Yang, Caroline Parins-Fukuchi, and Kathy Kron for helpful discussions and Luke Harmon, Eric Roalson, Matt Pennell, Alexis Stamatakis, and an anonymous reviewer for valuable feedback on drafts. Computations were performed on the Wake Forest University DEAC Cluster, a centrally managed resource with support provided in part by the University.

## <h1>FUNDING INFORMATION

S.A.S. and J.W.B. were supported by National Science Foundation Assembling, Visualizing, and Analyzing the Tree of Life Grant 1208809.

## <h1>LITERATURE CITED

- Aberer, A. J., K. Kobert, and A. Stamatakis. 2014. ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution* 31: 2553–2556.
- Aberer, A. J., D. Krompass, and A. Stamatakis. 2012. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Systematic Biology* 62: 162–166.
- Alfaro, M. E., S. Zoller, and F. Lutzoni. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution* 20: 255–266.
- Allman, E. S., L. S. Kubatko, and J. A. Rhodes. 2017. Split scores: a tool to quantify phylogenetic signal in genome-scale data. *Systematic Biology* 66: 620–636.
- Allman, E. S., and J. A. Rhodes. 2004. Quartets and parameter recovery for the general Markov model of sequence mutation. *Applied Mathematics Research Express* 2004: 107–131.
- Anderson, C. N. K., L. Liu, D. Pearl, and S. V. Edwards. 2012. Tangled trees: the challenge of inferring species trees from coalescent and noncoalescent genes. In M. Anisimova [ed.], *Evolutionary genomics, statistical and computational methods*, vol. 2, 3–28. Humana Press, New York, New York, USA.
- Andrews, D. W. K. 2000. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* 68: 399–405.
- Ané, C., B. Larget, D. A. Baum, S. D. Smith, and A. Rokas. 2006. Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution* 24: 412–426.

Banks, J. A., T. Nishiyama, M. Hasebe, J. L. Bowman, M. Gribskov, C. dePamphilis, V. A. Albert, et al. 2011. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332: 960–963.

Barrett, C. F., W. J. Baker, J. R. Comer, J. G. Conran, S. C. Lahmeyer, J. H. Leebens-Mack, J. Li, et al. 2015. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytologist* 209: 855–870.

Beaulieu, J. M., D. C. Tank, and M. J. Donoghue. 2013. A Southern Hemisphere origin for campanulid angiosperms, with traces of the break-up of Gondwana. *BMC Evolutionary Biology* 13: 80.

Bowe, L. M., G. Coat, and C. W. dePamphilis. 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceedings of the National Academy of Sciences, USA* 97: 4092–4097.

Brown, J. M., and A. R. Lemmon. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Systematic Biology* 56: 643–655.

Brown, J. M. and R. C. Thomson. 2017. Bayes factors unmask highly variable information content, bias, and extreme influence in phylogenomic analyses. *Systematic Biology* 66: 517–530.

Brown, J. W., J. F. Walker, and S. A. Smith. 2017. Phyx: phylogenetic tools for unix. *Bioinformatics* 33: 1886–1888.

Cannon, S. B., M. R. McKain, A. Harkess, M. N. Nelson, S. Dash, M. K. Deyholos, Y. Peng, et al. 2015a. Data from: Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.ff1tq>.

Cannon, S. B., M. R. McKain, A. Harkess, M. N. Nelson, S. Dash, M. K. Deyholos, Y. Peng, et al. 2015b. Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Molecular Biology and Evolution* 32: 193–210.

Chang, Y., and S. W. Graham. 2011. Inferring the higher-order phylogeny of mosses (Bryophyta) and relatives using a large, multigene plastid data set. *American Journal of Botany* 98: 839–849.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, et al. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528.



- Chaw, S. M., A. Zharkikh, H. M. Sung, T. C. Lau, and W. H. Li. 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. *Molecular Biology and Evolution* 14: 56–68.
- Chifman, J., and L. Kubatko. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30: 3317–3324.
- Crane, P. R. 1985. Phylogenetic analysis of seed plants and the origin of angiosperms. *Annals of the Missouri Botanical Garden* 72: 716–793.
- Cummings, M. P., S. A. Handley, D. S. Myers, D. L. Reed, A. Rokas, and K. Winka. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. *Systematic Biology* 52: 477–487.
- Degnan, J. H., and L. A. Salter. 2005. Gene tree distributions under the coalescent process. *Evolution* 59: 24–37.
- Douady, C. J., F. Delsuc, Y. Boucher, W. F. Doolittle, and E. J. P. Douzery. 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Molecular Biology and Evolution* 20: 248–254.
- Driskell, A. C., C. Ané, J. G. Burleigh, M. M. McMahon, B. C. O'Meara, and M. J. Sanderson. 2004. Prospects for building the tree of life from large sequence databases. *Science* 306: 1172–1174.
- Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.
- Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution* 28: 2239–2252.
- Eaton, D. A. R., E. L. Spriggs, B. Park, and M. J. Donoghue. 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Systematic Biology* 66: 399–412.
- Edwards, E. J., R. Nyffeler, and M. J. Donoghue. 2005. Basal cactus phylogeny: implications of *Pereskia* (Cactaceae) paraphyly for the transition to the cactus life form. *American Journal of Botany* 92: 1177–1188.
- Edwards, S. V., Z. Xi, A. Janke, B. C. Faircloth, J. E. McCormack, T. C. Glenn, B. Zhong, et al. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution* 94: 447–462.

- Efron, B. 1992. Bootstrap methods: another look at the jackknife. *In* S. Kotz and N. L. Johnson [eds.], *Breakthroughs in statistics*, vol. II, Methodology and distribution, 569–593. Springer, New York, New York, USA.
- Eklund, H., J. A. Doyle, and P. S. Herendeen. 2004. Morphological phylogenetic analysis of living and fossil Chloranthaceae. *International Journal of Plant Sciences* 165: 107–151.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791.
- Felsenstein, J., and H. Kishino. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. *Systematic Biology* 42: 193–200.
- Fletcher, W., and Z. Yang. 2009. INDELible: a flexible simulator of biological sequence evolution. *Molecular Biology and Evolution* 26: 1879–1888.
- Gaither, J., and L. Kubatko. 2016. Hypothesis tests for phylogenetic quartets, with applications to coalescent-based species tree inference. *Journal of Theoretical Biology* 408: 179–186.
- Givnish, T. J., M. Ames, J. R. McNeal, M. R. McKain, P. R. Steele, C. W. dePamphilis, S. W. Graham, et al. 2010. Assembling the tree of the monocotyledons: plastome sequence phylogeny and evolution of Poales. *Annals of the Missouri Botanical Garden* 97: 584–616.
- Givnish, T. J., A. Zuluaga, I. Marques, V. K. Y. Lam, M. S. Gomez, W. J. D. Iles, M. Ames, et al. 2016. Phylogenomics and historical biogeography of the monocot order Liliales: out of Australia and through Antarctica. *Cladistics* 32: 581–605.
- Goremykin, V. V., S. V. Nikiforova, D. Cavalieri, M. Pindo, and P. Lockhart. 2015. The root of flowering plants and total evidence. *Systematic Biology* 64: 879–891.
- Goulet, B. E., F. Roda, and R. Hopkins. 2017. Hybridization in plants: old ideas, new techniques. *Plant Physiology* 173: 65–78.
- Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328: 710–722.
- Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* 59: 307–321.
- Hahn, M. W., and L. Nakhleh. 2015. Irrational exuberance for resolved species trees. *Evolution* 70: 7–17.

- Hillis, D. M. and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* 42: 182–192.
- Hinchliff, C. E. and E. H. Roalson. 2013. Using supermatrices for phylogenetic inquiry: an example using the sedges. *Systematic Biology* 62: 205–219.
- Hinchliff, C. E., and S. A. Smith. 2014a. Data from: Some limitations of public sequence data for phylogenetic inference (in plants). Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.450qq>.
- Hinchliff, C. E., and S. A. Smith. 2014b. Some limitations of public sequence data for phylogenetic inference (in plants). *PLoS ONE* 9: e98986.
- Holder, M. T., J. Sukumaran, and P. O. Lewis. 2008. A justification for reporting the majority-rule consensus tree in Bayesian phylogenetics. *Systematic Biology* 57: 814–821.
- Hudson, R. R. 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* 18: 337–338.
- Huelsenbeck, J., and B. Rannala. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology* 53: 904–913.
- Jiao, Y., N. J. Wickett, S. Ayyampalayam, A. S. Chanderbali, L. Landherr, P. E. Ralph, L. P. Tomsho, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.
- Kobert, K., L. Salichos, A. Rokas, and A. Stamatakis. 2016. Computing the internode certainty and related measures from partial gene trees. *Molecular Biology and Evolution* 33: 1606–1617.
- Larget, B. 2013. The estimation of tree posterior probabilities using conditional clade probability distributions. *Systematic Biology* 62: 501–511.
- Larget, B., and D. L. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution* 16: 750–759.
- Lee, E. K., A. Cibrian-Jaramillo, S.-O. Kolokotronis, M. S. Katari, A. Stamatakis, M. Ott, J. C. Chiu, et al. 2011. A functional phylogenomic view of the seed plants. *PLoS Genetics* 7: e1002411.
- Matasci, N., L.-H. Hung, Z. Yan, E. J. Carpenter, N. J. Wickett, S. Mirarab, N. Nguyen, et al. 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience* 3: 17.

- McKain, M. R., H. Tang, J. R. McNeal, S. Ayyampalayam, J. I. Davis, C. W. dePamphilis, T. J. Givnish, et al. 2016. A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales. *Genome Biology and Evolution* 8: 1150–1164.
- Minh, B. Q., M. A. T. Nguyen, and A. von Haeseler. 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* 30: 1188–1195.
- Mirarab, S., R. Reaz, M. S. Bayzid, T. Zimmermann, M. S. Swenson, and T. Warnow. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30: i541–i548.
- Misof, B., B. Meyer, B. von Reumont, P. Kuck, K. Misof, and K. Meusemann. 2013. Selecting informative subsets of sparse supermatrices increases the chance to find correct trees. *BMC Bioinformatics* 14: 348.
- Nickrent, D. L., J. P. Der, and F. E. Anderson. 2005. Discovery of the photosynthetic relatives of the “Maltese mushroom” *Cynomorium*. *BMC Evolutionary Biology* 5: 38.
- Nickrent, D. L., C. L. Parkinson, J. D. Palmer, and R. J. Duff. 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Molecular Biology and Evolution* 17: 1885–1895.
- Nylander, J., F. Ronquist, J. Huelsenbeck, and J. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. *Systematic Biology* 53: 47–67.
- Ogden, T. H., and M. S. Rosenberg. 2006. Multiple sequence alignment accuracy and phylogenetic inference. *Systematic Biology* 55: 314–328.
- Palmer, J. D., D. E. Soltis, and M. W. Chase. 2004. The plant tree of life: an overview and some points of view. *American Journal of Botany* 91: 1437–1445.
- Pease, J. B., D. C. Haak, M. W. Hahn, and L. C. Moyle. 2016a. Data from: Phylogenomics reveals three sources of adaptive variation during a rapid radiation. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.182dv>.
- Pease, J. B., D. C. Haak, M. W. Hahn, and L. C. Moyle. 2016b. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biology* 14: e1002379.
- Pease, J. B., and M. W. Hahn. 2015. Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology* 64: 651–662.
- Pease, J. B., S. A. Smith, C. E. Hinchliff, J. W. Brown, and J. F. Walker. 2017. quartetsampling. Available at <https://github.com/fephyfofum/quartetsampling>.

- PPG I. 2016. A community-derived classification for extant lycophytes and ferns. *Journal of Systematics and Evolution* 54: 563–603.
- Pryer, K. M., L. Huiet, F.-W. Li, C. J. Rothfels, and E. Schuettpelz. 2016a. Data from: Maidenhair ferns, *Adiantum*, are indeed monophyletic and sister to shoestring ferns, vittarioids (Pteridaceae). Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.4m6s6>.
- Pryer, K. M., L. Huiet, F.-W. Li, C. J. Rothfels, and E. Schuettpelz. 2016b. Maidenhair ferns (*Adiantum*) are indeed monophyletic and sister to shoestring ferns, vittarioids (Pteridaceae). *Systematic Botany* 41: 17–23.
- Pryer, K. M., H. Schneider, A. R. Smith, R. Cranfill, P. G. Wolf, J. S. Hunt, and S. D. Sipes. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* 409: 618–622.
- Pryer, K. M., E. Schuettpelz, P. G. Wolf, H. Schneider, A. R. Smith, and R. Cranfill. 2004. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. *American Journal of Botany* 91: 1582–1598.
- Qiu, Y.-L., J. Lee, F. Bernasconi-Quadroni, D. E. Soltis, P. S. Soltis, M. Zanis, E. A. Zimmer, et al. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* 402: 404–407.
- Qiu, Y.-L., L. Li, B. Wang, Z. Chen, V. Knoop, M. Groth-Malonek, O. Dombrovskaya, et al. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences, USA* 103: 15511–15516.
- Ranwez, V., and O. Gascuel. 2001. Quartet-based phylogenetic inference: improvements and limits. *Molecular Biology and Evolution* 18: 1103–1116.
- Refulio-Rodriguez, N. F., and R. G. Olmstead. 2014. Phylogeny of Lamiidae. *American Journal of Botany* 101: 287–299.
- Reich, D., K. Thangaraj, N. Patterson, A. L. Price, and L. Singh. 2009. Reconstructing Indian population history. *Nature* 461: 489–494.
- Reid, N. M., S. M. Hird, J. M. Brown, T. A. Pelletier, J. D. McVay, J. D. Satler, and B. C. Carstens. 2013. Poor fit to the multispecies coalescent is widely detectable in empirical data. *Systematic Biology* 63: 322–333.

- Renzaglia, K. S., R. J. Duff, D. L. Nickrent, and D. J. Garbary. 2000. Vegetative and reproductive innovations of early land plants: implications for a unified phylogeny. *Philosophical Transactions of the Royal Society, B, Biological Sciences* 355: 769–793.
- Rice, D. W., A. J. Alverson, A. O. Richardson, G. J. Young, M. V. Sanchez-Puerta, J. Munzinger, K. Barry, et al. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* 342: 1468–1473.
- Rokas, A., B. L. Williams, N. King, and S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425: 798–804.
- Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Hohna, B. Larget, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542.
- Rosenberg, N. A. 2002. The probability of topological concordance of gene trees and species trees. *Theoretical Population Biology* 61: 225–247.
- Roure, B., D. Baurain, and H. Philippe. 2012. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution* 30: 197–214.
- Salichos, L., A. Stamatakis, and A. Rokas. 2014. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. *Molecular Biology and Evolution* 31: 1261–1271.
- Sanderson, M. J. 1995. Objections to bootstrapping phylogenies: a critique. *Systematic Biology* 44: 299–320.
- Sayyari, E., and S. Mirarab. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* 33: 1654–1668.
- Shavit Grievink, L., D. Penny, M. D. Hendy, and B. R. Holland. 2010. Phylogenetic tree reconstruction accuracy and model fit when proportions of variable sites change across the tree. *Systematic Biology* 59: 288–297.
- Shaw, A. J., P. Szövényi, and B. Shaw. 2011. Bryophyte diversity and evolution: windows into the early evolution of land plants. *American Journal of Botany* 98: 352–369.
- Shen, X.-X., C. T. Hittinger, and A. Rokas. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecology & Evolution* 1: 0126.

- Simmons, M. P. 2016. Mutually exclusive phylogenomic inferences at the root of the angiosperms: *Amborella* is supported as sister and Observed Variability is biased. *Cladistics* 33: 448–512.
- Simmons, M. P., and A. P. Norton. 2014. Divergent maximum-likelihood-branch-support values for polytomies. *Molecular Phylogenetics and Evolution* 73: 87–96.
- Smith, S. A., J. M. Beaulieu, and M. J. Donoghue. 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. *BMC Evolutionary Biology* 9: 37.
- Smith, S. A., J. M. Beaulieu, A. Stamatakis, and M. J. Donoghue. 2011. Understanding angiosperm diversification using small and large phylogenetic trees. *American Journal of Botany* 98: 404–414.
- Smith, S. A., M. J. Moore, J. W. Brown, and Y. Yang. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* 15: 150.
- Soltis, D. E., S. A. Smith, N. Cellinese, K. J. Wurdack, D. C. Tank, S. F. Brockington, N. F. Refulio-Rodriguez, et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* 98: 704–730.
- Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.
- Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Systematic Biology* 57: 758–771.
- Steel, M., and M. J. Sanderson. 2010. Characterizing phylogenetically decisive taxon coverage. *Applied Mathematics Letters* 23: 82–86.
- Steemans, P., A. L. Herisse, J. Melvin, M. A. Miller, F. Paris, J. Verniers, and C. H. Wellman. 2009. Origin and radiation of the earliest vascular land plants. *Science* 324: 353–353.
- Stenz, N. W., B. Larget, D. A. Baum, and C. Ané. 2015. Exploring tree-like and non-tree-like patterns using genome sequences: an example using the inbreeding plant species *Arabidopsis thaliana* (L.) Heynh. *Systematic Biology* 64: 809–823.
- Strimmer, K., N. Goldman, and A. von Haeseler. 1997. Bayesian probabilities and quartet puzzling. *Molecular Biology and Evolution* 14: 210–211.

Strimmer, K., and A. von Haeseler. 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences, USA* 94: 6815–6819.

Suzuki, Y., G. V. Glazko, and M. Nei. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proceedings of the National Academy of Sciences, USA* 99: 16138–16143.

Swofford, D. L., and J. Sullivan. 2003. Phylogeny inference based on parsimony and other methods using PAUP\*. In P. Lemley, M. Salemi, and A.-M. Vandamme [eds.], *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*, 60–206. Cambridge University Press, Cambridge, UK.

Taylor, Z. N., D. W. Rice, and J. D. Palmer. 2015. The complete moss mitochondrial genome in the angiosperm *Amborella* is a chimera derived from two moss whole-genome transfers. *PLoS ONE* 10: e0137532.

Testo, W. and M. Sundue. 2016. A 4000-species dataset provides new insight into the evolution of ferns. *Molecular Phylogenetics and Evolution* 105: 200–211.

The Amborella Genome Project. 2013. The *Amborella* genome and the evolution of flowering plants. *Science* 342: 1241089–1241089.

Walker, J. F., J. W. Brown, and S. A. Smith. 2017a. Site and gene-wise likelihoods unmask influential outliers in phylogenomic analyses. *bioRxiv*, <https://doi.org/10.1101/115774>.

Walker, J. F., Y. Yang, M. J. Moore, J. Mikenas, A. Timoneda, S. F. Brockington, and S. A. Smith. 2017b. Data from: Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.vn730>.

Walker, J. F., Y. Yang, M. J. Moore, J. Mikenas, A. Timoneda, S. F. Brockington, and S. A. Smith. 2017c. Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. *American Journal of Botany* 104: 858–867.

Washburn, J. D., J. C. Schnable, G. Davidse, and J. C. Pires. 2015. Phylogeny and photosynthesis of the grass tribe Paniceae. *American Journal of Botany* 102: 1493–1505.

Wickett, N. J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences, USA* 111: E4859–E4868.



- Wiens, J. J. and M. C. Morrill. 2011. Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Systematic Biology* 60: 719–731.
- Wilcox, T. P., D. J. Zwickl, T. A. Heath, and D. M. Hillis. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Molecular Phylogenetics and Evolution* 25: 361–371.
- Wilkinson, M. 1996. Majority-rule reduced consensus trees and their use in bootstrapping. *Molecular Biology and Evolution* 13: 437–444.
- Xi, Z., L. Liu, J. S. Rest, and C. C. Davis. 2014a. Coalescent versus concatenation methods and the placement of *Amborella* as sister to water lilies. *Systematic Biology* 63: 919–932.
- Xi, Z., L. Liu, J. S. Rest, and C. C. Davis. 2014b. Data from: Coalescent versus concatenation methods and the placement of *Amborella* as sister to water lilies. Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.qb251>.
- Xi, Z., J. S. Rest, and C. C. Davis. 2013. Phylogenomics and coalescent analyses resolve extant seed plant relationships. *PLoS ONE* 8: e80870.
- Yang, Y., M. J. Moore, S. F. Brockington, D. E. Soltis, G. K.-S. Wong, E. J. Carpenter, Y. Zhang, et al. 2015. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Molecular Biology and Evolution* 32: 2001–2014.
- Zanne, A. E., D. C. Tank, W. K. Cornwell, J. M. Eastman, S. A. Smith, R. G. FitzJohn, D. J. McGlinn, et al. 2014a. Data from: Three keys to the radiation of angiosperms into freezing environments. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.63q27.2>.
- Zanne, A. E., D. C. Tank, W. K. Cornwell, J. M. Eastman, S. A. Smith, R. G. FitzJohn, D. J. McGlinn, et al. 2014b. Three keys to the radiation of angiosperms into freezing environments. *Nature* 506: 89–92.
- Zhang, Z.-H., L. I. Chun-Qi, and L. I. Jianhua. 2009. Phylogenetic placement of *Cynomorium* in Rosales inferred from sequences of the inverted repeat region of the chloroplast genome. *Journal of Systematics and Evolution* 47: 297–304.
- Zhong, B., L. Liu, Z. Yan, and D. Penny. 2013a. Origin of land plants using the multispecies coalescent model. *Trends in Plant Science* 18: 492–495.
- Zhong, B., Z. Xi, V. V. Goremykin, R. Fong, P. A. Mclenachan, P. M. Novis, C. C. Davis, et al. 2013b. Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes. *Molecular Biology and Evolution* 31: 177–183.

Zhou, X., S. Lutteropp, L. Czech, A. Stamatakis, M. von Looz, and A. Rokas. 2017. Quartet-based computations of internode certainty provide accurate and robust measures of phylogenetic incongruence. *bioRxiv*, <https://doi.org/10.1101/168526>.

Zwickl, D. J. and D. M. Hillis. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* 51: 588–598.

Zwickl, D. J., J. C. Stein, R. A. Wing, D. Ware, and M. J. Sanderson. 2014. Disentangling methodological and biological sources of gene tree discordance on *Oryza* (Poaceae) chromosome 3. *Systematic Biology* 63: 645–659.

**TABLE 1.** Quartet Sampling (QS) score interpretation.

Example QS score (QC/QD/QI)	Interpretation
1.0/–/1.0	Full support: All sampled quartet replicates support the focal branch (QC = 1) with all trees informative when likelihood cutoffs are used (QI = 1).
0.5/0.98/0.97	Strong support: A strong majority of quartets support the focal branch (QC = 0.5), and the low skew in discordant frequencies (QD $\approx$ 1) indicate no alternative history is favored.
0.7/0.1/0.97	Strong support with discordant skew: A strong majority of quartets support the focal branch (QC = 0.7), but the skew in discordance (QD = 0.1) indicates the possible presence of a supported secondary evolutionary history.
0.05/0.96/0.97	Weak support: Only a weak majority of quartets support the focal branch (QC = 0.05), and the frequency of all three possible topologies is similar (QD $\approx$ 1).
0.1/0.1/0.97	Weak support with discordant skew: Only a weak majority of quartets support the focal branch (QC = 0.1), and the skew in discordance (QD = 0.1) indicates the possible presence of a supported secondary evolutionary history.
–0.5/0.1/0.93	Counter-support: A strong majority of quartets support one of the alternative discordant quartet arrangement history (QC < 0; QD expected to be low).
1/0.97/0.05	Poorly informed: Despite supportive QC/QD values, only 5% of quartets passed the likelihood cutoff (QI = 0.05), likely indicating few informative sites.
0.0/0.0/1.0	Perfectly conflicted: The (unlikely) case where the frequencies of all three possible trees are equal and all trees are informative, which indicates a rapid radiation or highly complex conflict.

Notes: QC = Quartet Concordance; QD = Quartet Differential; QI = Quartet Informativeness.

**TABLE 2.** Quartet Sampling scores for key branches in the plant tree of life.

Branch	Quartet Concordance				Consensus interpretation
	HS2014	ZN2014	WI2014	XI2014	
Embryophytes (land plants)	0.35	n.t.	1.0	n.t.	strong support
Tracheophytes (vascular plants)	0.14	0.31	0.29	n.t.	moderate-strong support
Euphyllophytes (ferns + seed plants)	0.02	−0.06	0.44	n.t.	low/variable support
Spermatophytes (seed plants)	0.23	0.36	0.95	n.t.	strong support
Acrogymnospermae (gymnosperms)	0.37	0.32	0.92	1.0	strong support
Anthocerotophyta (hornworts)	0.54	0.94	1.0	n.t.	strong support
Bryophyta (mosses)	0.41	0.15	1.0	n.t.	moderate-strong support
Lycopodiophyta	0.38	0.32	0.89	n.t.	strong support
Magnoliophyta (angiosperms)	0.68	0.75	0.92	0.95	strong support
Marchantiophyta (liverworts)	0.15	0.8	1.0	n.t.	moderate-strong support
Polypodopsida (ferns)	0.23	0.46	1.0	n.t.	moderate support
Chloranthaceae + core angiosperms	−0.26	−0.04	n.m.	n.t.	counter-supported
Chloranthaceae + eudicots	n.m.	n.m.	−0.47	n.t.	counter-supported
Magnoliids	0.20	0.31	0.53	0.54	strong support
Eudicots	0.16	0.43	0.32	0.71	moderate-strong support
Asterids	−0.01	0.32	0.63	0.0	low/variable support
Rosids	0.05	n.m.	1.0	0.25	low/variable support
Monocots (including <i>Acorus</i> )	0.04	0.06	0.38	n.t.	low/variable support
Monocots (excluding <i>Acorus</i> )	0.01	−0.05	0.39	0.76	low/variable support
Liliales + commelinids	0.03	n.m.	n.t.	n.t.	low support
Liliales + Asparagales	n.m.	0.03	n.t.	n.t.	low support

Notes: n.t. = not testable with this data set; n.m. = not monophyletic in this tree.

**FIGURE 1.** Description of the Quartet Sampling method. (A) The focal branch “*b*” divides the phylogeny into four subclades  $\{S_1, S_2, S_3, S_4\}$  from which tips (A–J) are sampled. Two replicates with different sampled tips for the given branch are shown with the three possible unrooted topologies (one concordant and two discordant). (B) Each internal branch is labeled with a set of three scores (Quartet Concordance/Quartet Differential/Quartet Informativeness), which offer different, but complementary, information. Terminal branches are evaluated by the Quartet

Fidelity score, which reports the frequency of a taxon's inclusion in concordant topologies. (See Materials and Methods for full details and Appendix S1 for a technical description.)

**FIGURE 2.** Results of simulation testing of the Quartet Sampling (QS) method. (A) Quartet Concordance (QC) values converge on a central value with increasing numbers of replicates from randomly selected branches from simulated trees with 50, 100, and 500 taxa. (B) Mean QC (solid diamond) and Quartet Differential (QD; open circle) values with 5th–95th percentile (whiskers) across 100 replicates for branches QS16 (left) and QS12 (middle) for a simulated tree (Appendix S5) where the tree branch lengths were scaled by the factors on the  $x$ -axis (i.e., 1 is the original tree). As expected, shorter branch lengths will increase the level of incomplete lineage sorting and thus lower the QC scores. The right panel shows branch QS11 from the simulated tree with increasing levels of introgression introduced by simulation. As expected, QC and QD values decrease with increasing introgression. (C) Distributions of QC, Quartet Informativeness, and Quartet Fidelity values for HS2014 (black), ZN2014 (dotted black), and XI2014/CN2015/PR2016/WA2017 (similar distributions; gray solid). (D) Mean QC values (diamond) with 5th to 95th percentile (whiskers) for branches in HS2015 binned by the number of subtending taxa (i.e., moving rootward in the tree left to right). Overall mean is shown with horizontal dotted line.

**FIGURE 3.** Phylogeny from Hinchliff and Smith (2014b). (A) Full phylogeny with heat map coloration of branches by Quartet Concordance (QC) scores for internal branches: dark green (QC > 0.2), light green ( $0.2 \geq \text{QC} > 0$ ), light orange ( $0 \geq \text{QC} \geq -0.05$ ), or dark orange (QC < -0.05). (B) QC/Quartet Differential (QD)/Quartet Informativeness (QI) scores (200 replicates of full alignment) for major plant groups and key orders within angiosperms. QC/QD/QI scores are shown. Group names are for the ancestral branch (i.e., the “stem” branch), or a single Quartet Fidelity score is shown for monotypic tips. Major subgroups groups are highlighted with vertical labels. (C) Quartet Sampling (QS) scores for monocots (excluding *Acorus*). (D, E, F) QS scores for rosids, Bryophyta, and gymnosperms. *Abbreviations:* Acro, Acrogymnospermae; ANA, ANA grade; Aru, Arundinoideae; Bry, Bryophyta, Chl, Chloridoideae; Dan, Danthonioideae; Mar, Marchantiophyta; Poly, Polypodopsida.

**FIGURE 4.** Phylogeny from Zanne et al. (2014b). (A) Full phylogeny with heat map coloration of branches by Quartet Concordance (QC) scores for internal branches using same color scheme as in Fig. 3. (B) QC/Quartet Differential/Quartet Informativeness scores (200 replicates of full alignment) for major plant groups and key orders within angiosperms, using color scheme in Fig. 3. (C) Quartet Sampling (QS) scores shown for monocots (except *Acorus*). (D) QS scores for asterids. (E) QS scores for fern lineages and (F) QS scores for gymnosperm lineages.

*Abbreviations:* Alseu, Alseuosmiaceae; ANA, ANA-grade angiosperms; Argo, Argophyllaceae; Aster, Asteraceae; Bory, Boryaceae; Caly, Calycanthaceae; Eriach, Eriachneae; Good, Goodeniaceae; gym, gymnosperms; Hypox, Hypoxidaceae; Isach, Isachneae; Phell, Phellinaceae; Poly, Polypodopsida.

**FIGURE 5.** Maximum likelihood phylogeny spanning *Viridiplantae* from Fig. 2 in Wickett et al. (2014) with Quartet Concordance (QC)/Quartet Differential/Quartet Informativeness scores for 200 replicates of the full alignment. Nodes are colored according to QC score using the same color scheme as in Fig. 3. Bootstrap values (italicized in square brackets) from Wickett et al. (2014) are shown for comparison. Missing Quartet Sampling or bootstrap values indicate a perfect score. The three taxa with the lowest Quartet Fidelity values are highlighted. Species names have been excluded or abbreviated when two congeners are included.

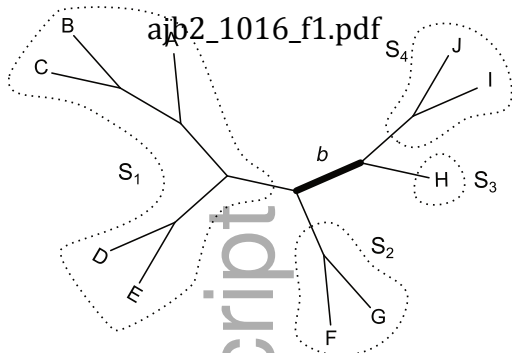
**FIGURE 6.** Key phylogenetic disagreements with Quartet Concordance (QC) scores for the same branches compared across various data sets. Branches for HS2014 and ZN2014 were resampled with 10,000 replicates. Branches for WI2014 and XI2014 were exhaustively sampled (>1000 replicates). Highlighting of QC values follows the same colors as in Fig. 3. “Conifers-II” refers to a hypothesized clade comprising the non-Pinales orders in Pinidae. *Abbreviations:* Gnet, Gnetidae; Pin, Pinidae.

**FIGURE 7.** Phylogeny of Pteridaceae ferns from Pryer et al. (2016b) with Quartet Concordance (QC)/Quartet Differential/Quartet Informativeness scores for 200 replicates of the full alignment. Nodes are colored according to QC score using the color scheme in Fig. 3. Bootstrap/Shimodaira–Hasegawa (SH) test/posterior probability values (italicized in square brackets) are

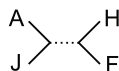
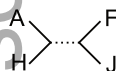
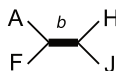
shown for comparison. Omitted values indicate a perfect score. The three taxa with the lowest Quartet Fidelity values are highlighted. *Abbreviations: Pityro, Pityrogramma.*

**FIGURE 8.** Quartet Sampling scores for phylogenies from whole-transcriptome data. Omitted values indicate a perfect score. Nodes are colored according to Quartet Concordance score using the same color scheme as in Fig. 3. (A) Phylogeny of *Solanum* sect. *Lycopersicon* from Pease et al. (2016b). Bootstrap values (italicized in square brackets) are shown for comparison. (B) Phylogeny of Caryophyllales from Walker et al. (2017c). Internode certainty scores (light grey) are shown for comparison (all bootstrap and Shimodaira–Hasegawa (SH) test scores were 100). The three taxa with the lowest Quartet Fidelity values are highlighted.

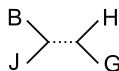
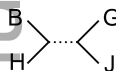
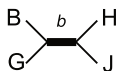
A



Replicate 1



Replicate 2

Concordant  
quartet topologyDiscordant  
topology 1Discordant  
topology 2

B

**Quartet Sampling Internal Node Scores** = 0.52 / 0.91 / 0.95

**Quartet Concordance (QC)**

How often is the concordant quartet inferred over both discordant quartets?

QC=1 → all concordant

QC=0 → equivocal conc./disc.

QC&lt;0 → discordant &gt; conc.

**Quartet Differential (QD)**

Are discordant #1 and #2 frequencies equal or skewed?

QD=1 → equal #1 and #2

QD=0.3 → skewed

QD=0 → all #1 or #2

**Quartet Informativeness (QI)**

What proportion of replicates were informative?  
(exceeded likelihood differential)

QI=1 → all informative

QI=0.3 → 30% informative

QI=0 → none informative

**Quartet Sampling Terminal Node Scores** = (0.52)

**Quartet Fidelity (QF)**

When this taxon is sampled, how often does it produce a concordant topology?

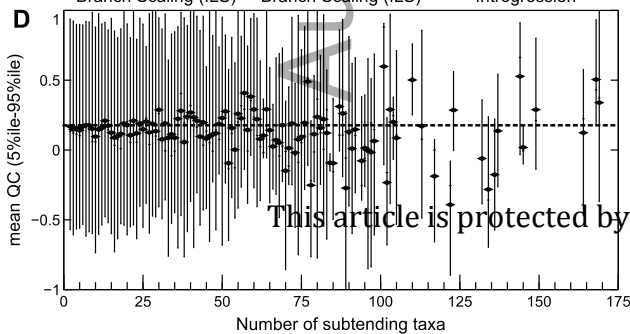
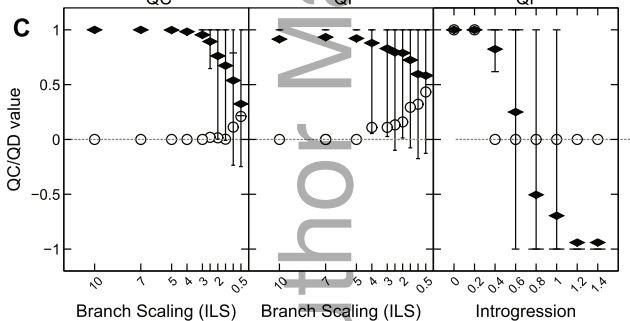
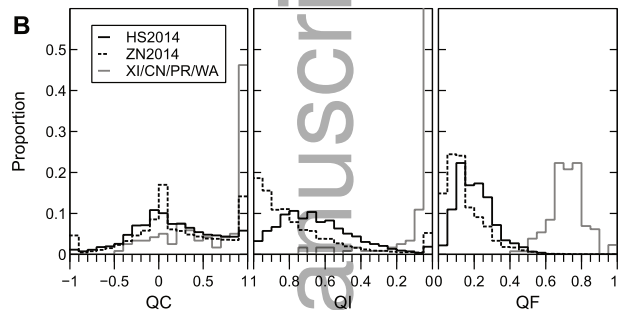
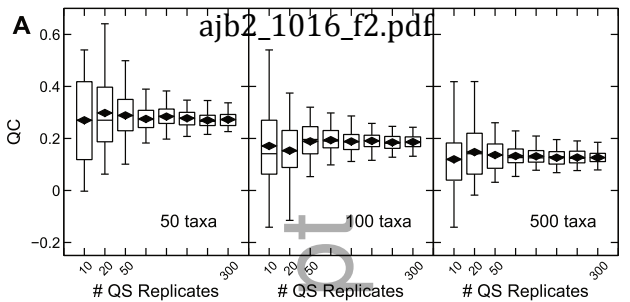
Examples:

QF=1 → all concordant

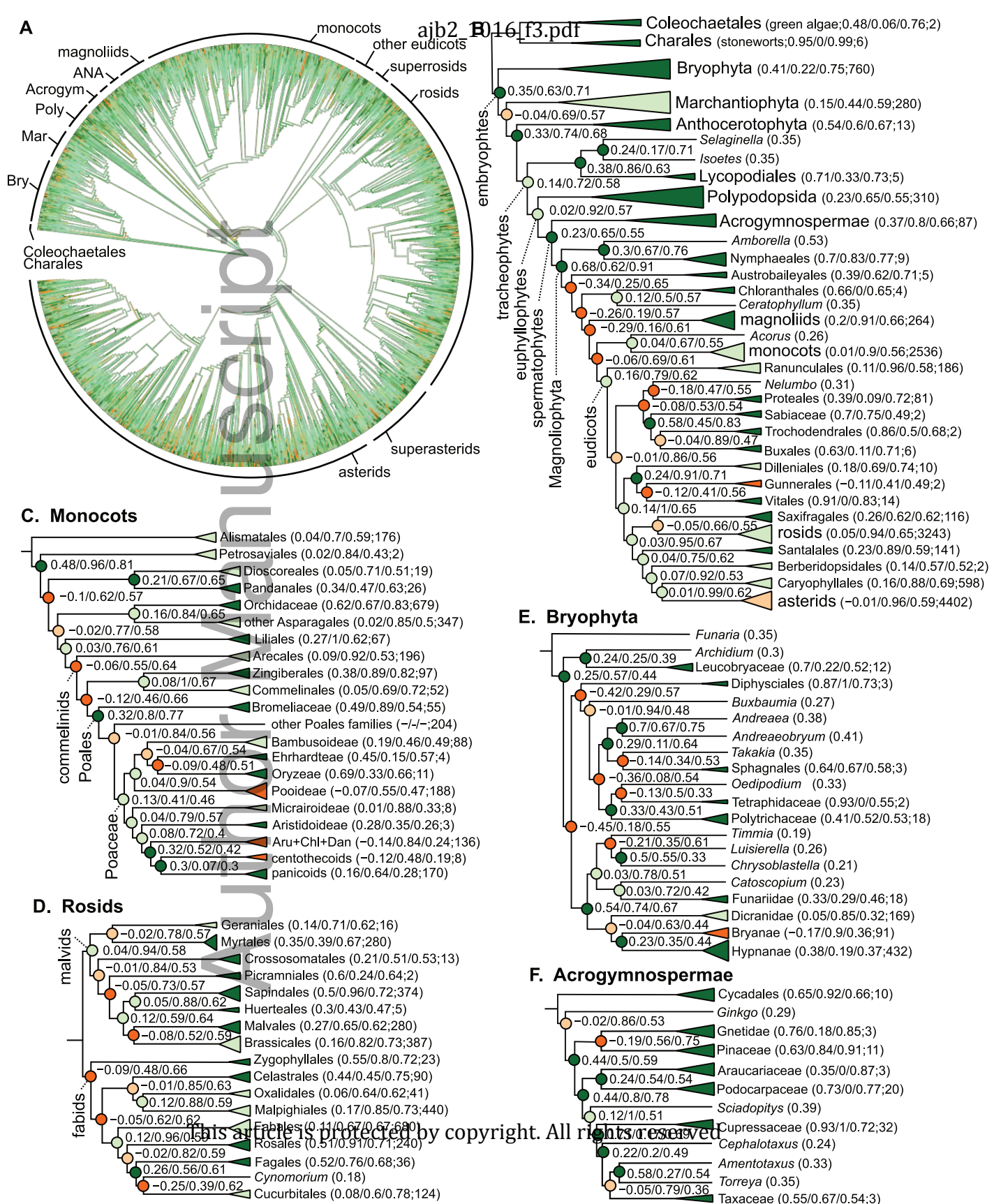
QF=0.1 → 10% concordant

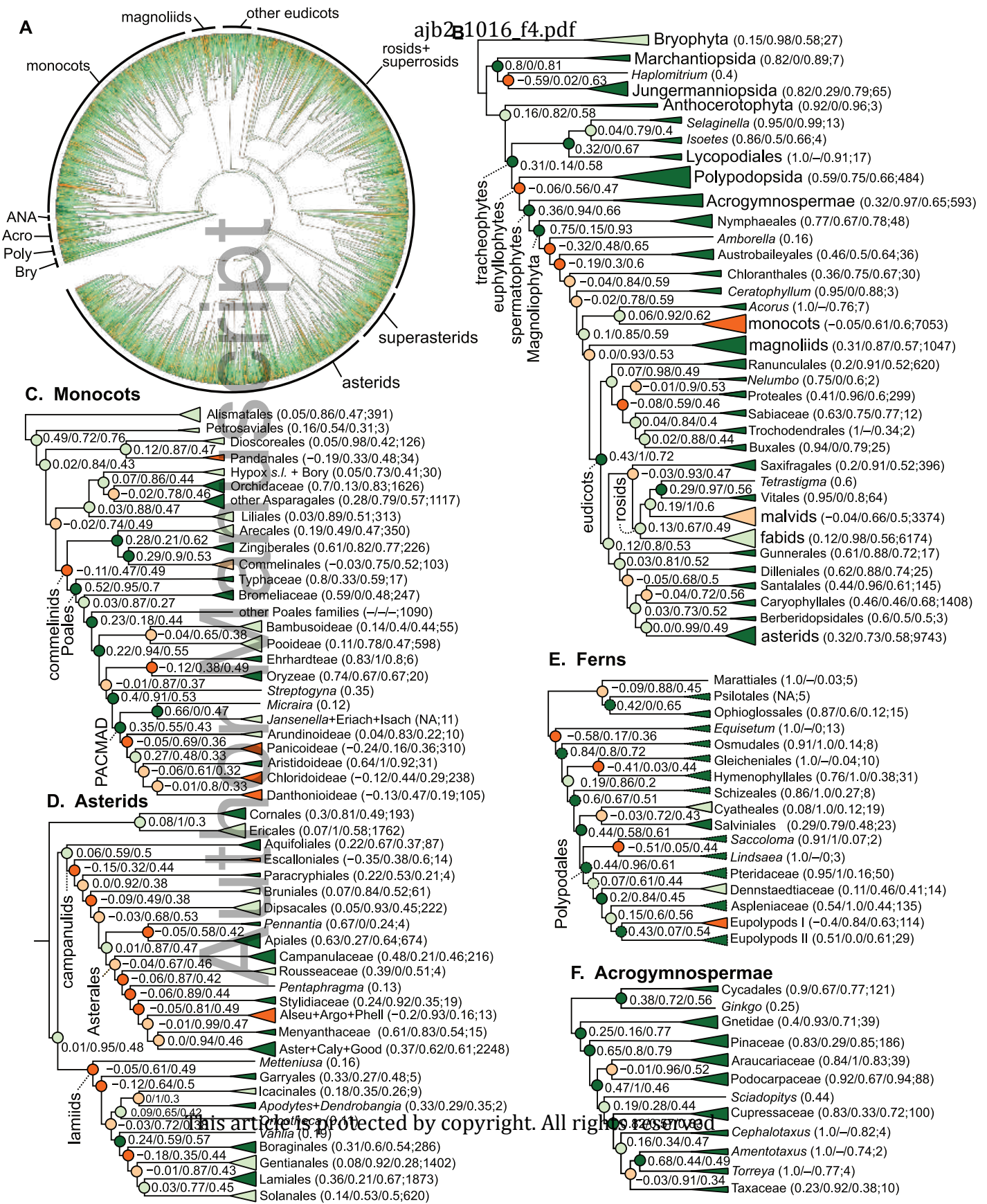
QF=0 → none concordant

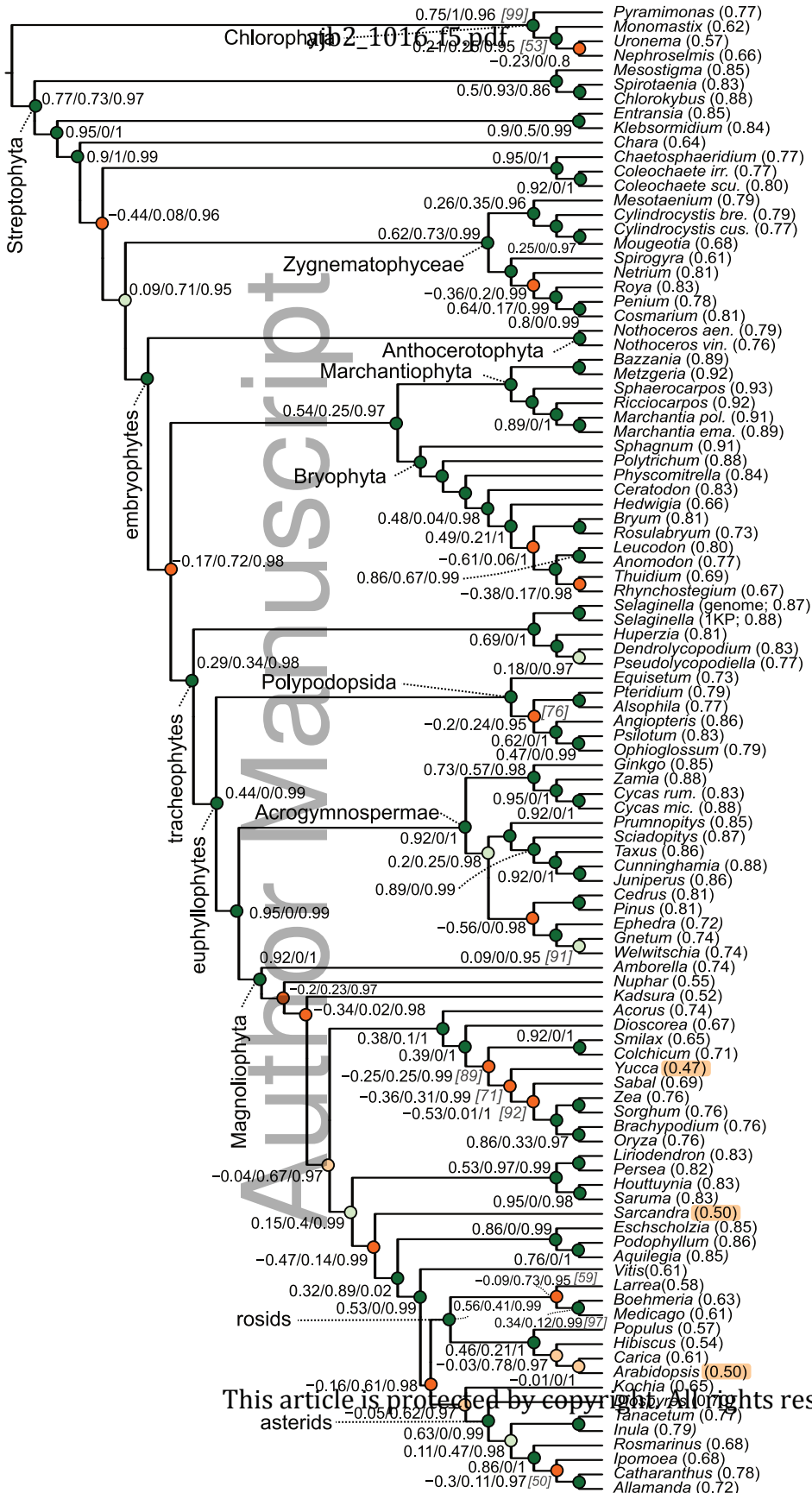
This article is protected by copyright













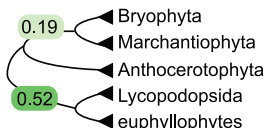
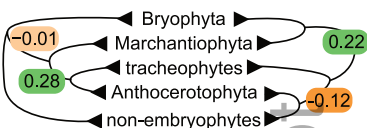
## A. non-vascular plants

HS2014

ajb2\_1016\_f6.pdf

WI2014

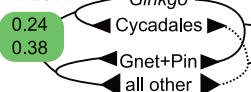
ZN2014



## B. *Ginkgo* + cycads

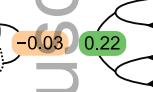
ZN2014

WI2014



HS2014

ZN2014



## C. Gnetidae + conifers

HS2014

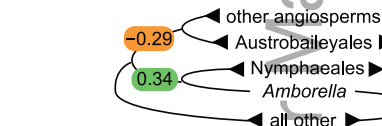
WI2014



## D. “ANA grade” angiosperms

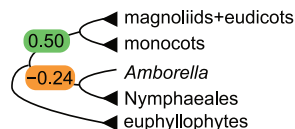
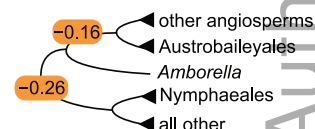
HS2014 (“*Ambo+Nym-first*”)

WI2014 (“*Amborella-first*”)

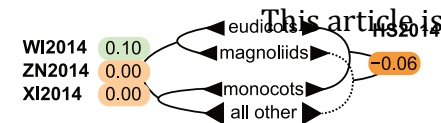


ZN2014 (“*Nymphaeales first*”)

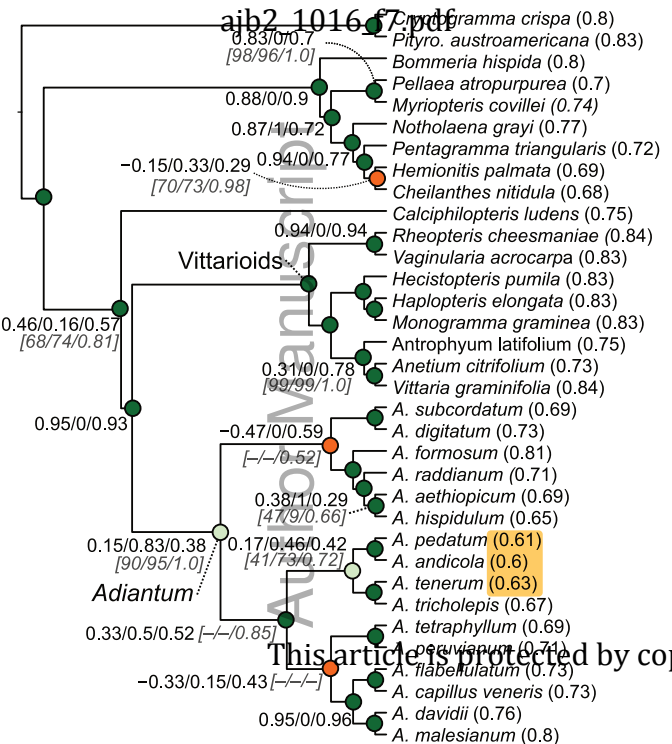
XI2014 (“*Ambo+Nym-first*”)



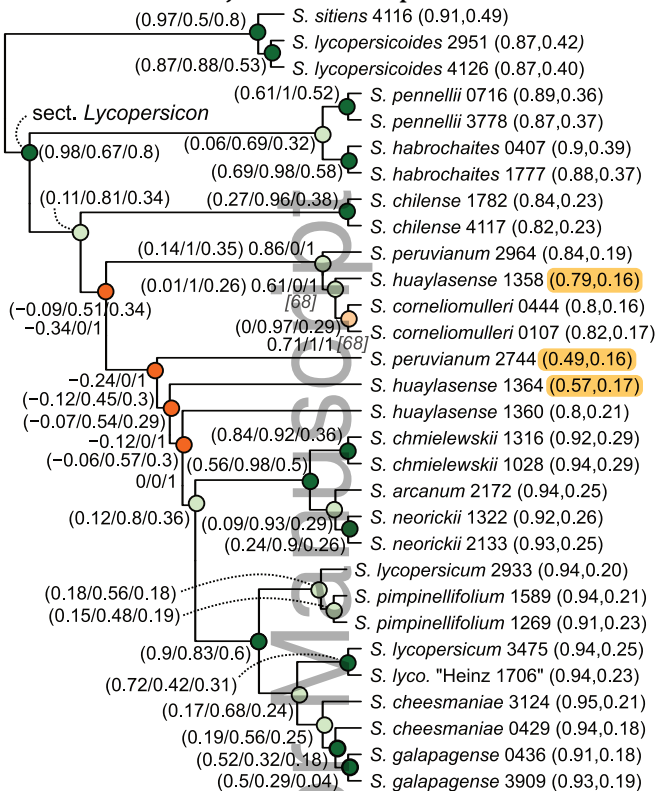
## E. magnoliids, monocots, and eudicots



This article is protected by copyright



## A. *Solanum* sect. *Lycopersicon*



## B. Caryophyllales

